Research and Diagnostic Validity of Whole Exome Sequencing in Neuromuscular Disease

Linran Zhou

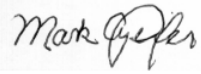
A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the Senior Honors Thesis in the Department of Biology in the College of Arts and Sciences.

Chapel Hill
2015

Approved:

Thesis Advisor

Reader



Reader

ABSTRACT

Linran Zhou: Research and Diagnostic Validity of Whole Exome Sequencing in Neuromuscular Disease
(Under the direction of Jonathan S. Berg)

Neuromuscular diseases (NMDs) are a broad group of inherited genetic disorders with a total global prevalence exceeding 1 in 3,000. Affecting the muscle and peripheral nervous system, NMDs result in significant disability and despite increased understanding of their molecular basis, the cause is unknown in the majority of patients. In an NCGENES patient cohort, we examined the diagnostic yield of whole exome sequencing (WES) for variants in genes currently associated with NMDs, neuropathies, and myopathies, and we evaluated the literature for known information about these genes. We analyzed the variants from WES in genes not on the current diagnostic lists to search for novel, rare, deleterious variants in possible disease candidate genes and identify patient candidates for further molecular analysis. Based on literature and clinical reports, we also evaluated the construction of the current diagnostic lists, expanding the myopathy and neuropathy lists by over 80 genes each. The lists were curated so that NMD list encompassed all myopathy and neuropathy genes, while the myopathy and neuropathy lists remained as unique as possible. Based on our overall analysis, we constructed a schematic to aid clinicians in deciding when to apply WES for patients with a neuromuscular condition of suspected genetic etiology.

ACKNOWLEDGEMENTS

I offer my deepest thanks to Dr. Jonathan Berg, Dr. Gloria Haskell, and Dr. Laura Milko for mentoring me, not only through this project, but through all my time at the Berg lab. Truly, you have been nothing but kind, patient, and motivating to me, and words cannot do justice the gratitude that I feel. My experience at UNC would not have been the same with you. I can only hope this project serves as a respectful tribute and dedication to all that you have done for me.

I cannot forget everyone else I have met at the Berg lab. Thank you, Daniel, Michael, Christian, Kristen, and Lacey, for all of your kind words, encouragement, and help. It has meant more to me that you would know.

Thank you, Dr. Peifer, for your sponsorship and continued advocacy and interest in my project. Your meticulous eye and astute questioning were paramount in ensuring I explained my process clearly and effectively.

Thank you, Dr. Bruno, for all of your work with this class. Through this experience, I have learned a lot about my own abilities, as well as shortcomings, and for that, I cannot thank you enough.

Finally, I must also thank the UNC Department of Biology and the Department of Genetics, for all the opportunities they have provided for me to further my growth and education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MAF	Minor allele frequency
NCGENES	North Carolina Clinical Genomic Evaluation by Next-Generation Exome Sequencing
NGS	Next-generation sequencing
NMDs	Neuromuscular disease (or disorder)
OMIM	Online Mendelian Inheritance in Man
VUS	Variant of unknown significance
WES	Whole exome sequencing

INTRODUCTION

Neuromuscular diseases (NMDs) are a broad category of genetic disorders that affect both the muscle and the peripheral nervous system. Consequently, they principally affect the ability to perform voluntary movements, leading to significant disability and morbidity in both children and adults.[1] NMDs consist of over 200 monogenic disorders, which include disorders like Duchenne muscular dystrophy (DMD), amyotrophic lateral sclerosis, limb-girdle muscular dystrophies (LGMDs) and hereditary motor and sensory neuropathies. Analysis of reported studies suggests that roughly 1 in 3,500 individuals across both sexes in the global population could be expected to present with an inherited NMD in childhood or later life. If rarer NMDs are also included in the analysis, it is estimated that the overall global prevalence could exceed 1 in 3,000, highlighting the need for research and development in this area.[1,2] NMDs range in time of onset from *in utero* to old age, but to a large extent, they first present in infancy, childhood, or adolescence.[3]

Over the past 25 years, there have been monumental research and clinical endeavors that have revealed much about the molecular bases of NMDs, catalyzed by the identification of many variants in genes and loci in people with the disorders and significant improvements in the treatment of symptoms and complications.[1,3] Treatments such as assisted ventilation, physiotherapy, and orthotics have greatly increased the life expectancy and quality of life for many patients with NMDs.[3] There are databases that catalogue the known genes and linked loci with genes not yet identified associated with NMDs (summarized in Figure 1.)

Table 1. Categories of Neuromuscular Disorders.

Disease category	Number of genes/loci
Muscular dystrophies	31
Congenital muscular dystrophies	25
Congenital myopathies	30
Distal myopathies	15
Other myopathies	22
Myotonic syndromes	9
Ion channel muscle diseases	19
Malignant hyperthermias	6
Metabolic myopathies	23
Hereditary cardiomyopathies	102
Congenital myasthenic syndromes	21
Spinal muscular atrophies (including motor neuron diseases)	34
Hereditary ataxias	57
Hereditary motor and sensory neuropathies	60
Hereditary paraplegias	41
Other neuromuscular disorders	24
Total	519

Figure 1.[3] Categories of neuromuscular disorders, summarized from a November 2011 table published by the journal *Neuromuscular Disorders*. There are 16 categories of neuromuscular disorders. Taken from Laing, 2012.

Despite this progress and mounting clinical and molecular evidence, the causes of NMDs remain unknown in a majority of patients. Major challenges that persist in attempting to link genetic variants to specific phenotypes in affected families or individuals include lack of segregation data, especially in sporadic cases, non-specific clinical features that could be indicative of multiple disorders, and genetic and phenotypic heterogeneity. Some genes associated with NMDs, such as TTN, DMD, or RYR, are among the largest human genes.[1] Combined with the sheer variety of variants in those known genes – point mutations and small indels, deletions of entire genes, deletion duplication of exons, among others – and the effects of unknown genes potentially associated with NMDs, these factors greatly complicate molecular and genetic analysis.[1]

Recently, the use of next generation sequencing (NGS) techniques has become a popular and effective technique for massively parallel analysis of many genes and loci. These technologies have led to successful identification of several Mendelian disease genes, some of

which were not amenable to linkage analysis as a result of disease rarity and strong negative selection on causal mutations, leading to those mutations often arising from *de novo* events.[1,4] One key application of NGS technologies is whole exome sequencing (WES).

The human genome consists of about 3.2 billion nucleotides and has about 23,500 genes. The protein-coding region of each gene is referred to as an exon, and the human genome contains about 180,000 exons. These exons are collectively referred to as the exome, which makes up about 1% of the human genome and represents the totality of all protein-coding regions. Although the exome does not include roughly 99% of the human genome, about three-quarters of currently known pathogenic variants affect exons, so much of the focus in medically-applicable sequencing is targeted towards the exome.[5] Exome sequencing has demonstrated its value in many notable cases, such as when it was used to identify variants in *DHODH* as the cause of Miller disorder, a rare Mendelian disorder of previously unknown etiology.[6] While thorough technical descriptions of exome sequencing can be found elsewhere[7,8], the general process is as follows. First, DNA is extracted from white blood cells, sheared into fragments, hybridized to mRNA baits to enrich for exomic regions, and sequenced using various sequencing technologies. The NGS sequencing experiments quickly produce millions to billions of short sequence reads, which are aligned to the human reference genome. After mapping, a series of quality checks and processing steps are applied to minimize sequencing artifacts that could interfere with variant calling. Finally, variants in the sequenced samples are detected, annotated with useful biological information, and filtered to output a manageable list of candidate variants for experimental verification.[9,10] As with any test, WES is not immune to false-positive and false-negative results, so the generated data must be carefully analyzed to be medically applicable.

This study is part of the larger, comprehensive North Carolina Clinical Genomic Evaluation by Next-Generation Exome Sequencing (NCGENES) project. As the costs of exome and genome sequencing decrease, there is increasing interest in clinical genome-scale analysis, and as a result, research is needed to understand the best practices. The four overarching aims of NCGENES are as follows: identify how WES performs as a diagnostic tool, learn how incidental findings affect patient choices and healthcare decisions, develop a practical and ethical clinical framework to incorporate WES data, and bring WES to medically-underserved groups to explore opportunities and challenges. Thus, the questions and objectives of this study reflected the overall aims of NCGENES. First, given the current diagnostic lists[I] for NMDs, myopathy, and neuropathy, we were interested in the diagnostic yield[II] for each of those lists. Since these disease categories are often phenotypically variable, many patients were run on more than one diagnostic list, which provided an opportunity to explore relationships between the clinical features and histories patients presented with, what lists they were run on, and which lists they tested positive for, if any. This information, combined with literature analysis, helped determine how the lists could be improved for clinical use, not only by adding new genes to each list, but also adjusting how the lists are structured and designed, which is important in determining which ones to use for a particular patient. We hypothesized that there were distinguishing clinical features that would help clinicians and researchers predict from which lists patients would most likely return positive results. The patients who did not return variants in known disease genes presented an opportunity for gene discovery – identifying rare, deleterious variants in new genes that could explain the disease phenotype. The ultimate goal of the study is to explore and

[I] A list of genes that have been found to be associated with that category of disease that also includes the pattern of inheritance and a tier based on the number of unrelated individuals in which disease-associated variants in that gene have been found

[II] The likelihood that a test will provide the information needed to reach a diagnosis (i.e.: the proportion of patients run on each diagnostic list who received a (likely) positive result)

improve on how WES is used in NMDs, so the analysis of the positive results, curation of the lists, and research sweeps in the negative results was distilled into a framework to aid clinicians in determining when to use WES in patients with suspected NMDs.

MATERIALS AND METHODS

Whole Exome Sequencing and Variant Calling

A total of 83 patients were enrolled into the NCGENES neuromuscular, myopathy, or neuropathy cohorts after meetings with physicians and genetics counselors, and each patient was run on at least one of the NMD, myopathy, or neuropathy diagnostic lists (Supplemental 1.1-1.2). The overall WES workflow and variant calling pipeline is illustrated in a schematic (Appendix 1.1). DNA was extracted from blood samples from the patients, and whole-exome sequencing was performed by NCGENES staff using the SureSelect^{XT} Target Enrichment System Kit for Illumina Multiplexed Sequencing (Appendix 1.2). Reads were stored and annotated in VarDb, an in-house variant database, based on variant frequency and predicted impact. An automated custom workflow developed by RENCI was used to return all variants in the 83 patients in genes on the diagnostic lists with a minor allele frequency (MAF) – the frequency at which the least abundant allele of a single nucleotide polymorphism (SNP) is present – in 1000 Genomes of less than five percent.

Analysis of Diagnostic Yield and Positive Results

WES results were pulled from the database for each patient who had been run on at least one of the NMD, myopathy, or neuropathy diagnostic lists. These results took the form of a list, in which each patient received a “positive” if they were found to have a (likely) pathogenic or deleterious variant or variants, “negative” if they were not, and “uncertain” if they were found to

have a variant of unknown significance (VUS). The category of "uncertain" comprises multiple possibilities. First, there are VUS in which it is uncertain whether the variant is indeed deleterious. The second possibility arises from VUS that could possibly contribute to disease but are not wholly consistent with the patient's phenotype. The third possibility is that the VUS is a heterozygous variant for a patient with a known recessive disorder, but a second variant was not found. Some patients received a "null" response, either indicating that their reported variants from WES had not yet been fully experimentally verified or that their findings were not yet recorded or updated. Diagnostic yields were calculated for the myopathy and neuropathy lists individually by dividing the number of patients who had received a positive result by the total number of patients who had been run on that list. For the patients who had been run on either both the myopathy and NMD lists or both the neuropathy and NMD lists, another diagnostic yield was calculated for the use of two lists in tandem (either myopathy and NMD or neuropathy and NMD) by dividing the number of patients who received a positive result on either list by the total number of patients who had been run on that combination of lists. Clinical information was pulled from REDCap for each patient who received a positive result from either combination of lists in order to compare the phenotype associated with the variant to the symptoms they displayed and their working clinical diagnosis.

Curation of the Diagnostic Lists

The initial NCGENES diagnostic lists for myopathy, neuropathy, and NMDs were compared to determine the extent to which the lists shared genes. From examining the literature and consulting with physicians, it was decided that the NMD list should contain all the genes from both the neuropathy and myopathy lists, in addition to any genes associated with an NMD that is not a neuropathy or myopathy. Using the clinical reports and literature from the Online

Mendelian Inheritance in Man (OMIM) database, information about each gene that was unique to one of the three lists was collected, including gene function, disease phenotypes, known variants associated with disease, and clinical data about the patients in which the variants were found. Each gene unique to the NMD list was evaluated using information from the OMIM database in order to determine if any of those genes could be additionally classified as a myopathy or neuropathy gene. A gene was added to either the myopathy or neuropathy list if there was clinical evidence showing at least one patient with an associated phenotype demonstrating characteristics of myopathy or neuropathy. To ensure that the diagnostic lists were up-to-date, the OMIM clinical synopses were queried for all disease phenotypes associated with either myopathy or neuropathy, and every resulting entry was examined to ensure that there was at least one patient with that phenotype who demonstrated signs of a neuropathy or myopathy, such as an idiosyncratic nerve conduction study or muscle biopsy result. All genes associated with a disease phenotype that met those criteria on OMIM were added to the respective list, and all new myopathy or neuropathy genes were also added to the NMD list. The updated diagnostic lists were reexamined to determine the level of overlap among those genes in order to verify that the structure and content of the new lists would be appropriate for future variant calling when analyzing WES reads.

Analysis of Negative Results

Using the initial NCGENES diagnostic lists, a research diagnostic (RDx) list of genes was generated from the *in silico* STRING database, which returned genes that encoded proteins that were known or predicted to interact with the proteins encoded by the myopathy, neuropathy, or NMD genes. This RDx list was used for gene discovery in the patients with negative diagnostic list results by computationally filtering their sequencing data and selecting for variants

in RDx genes. A frequency filter was applied to only obtain the variants in the patients with negative results with a MAF of less than one percent in the Exome Aggregation Consortium (ExAC) database, and the resulting variants were divided into truncating and missense variants. Each variant was manually curated in order to evaluate its potential pathogenicity. The parameters used to gauge pathogenicity were derived from reviewing literature examining best practices in characterizing WES variants and included the variant frequency across all NCGENES patients, the CADD scores, the MAF in ExAC, published gene and function information, and affected protein domains from UniProt. Clinical information from REDCap, such as family histories, symptoms, and previous testing, was used to assess inheritance patterns and narrow down the possible categories of NMDs. After manual curation of the negatives, the variants that passed each filter were collected for later confirmation and further characterization, and patients with interesting presentations and backgrounds were recorded for further, more in-depth analysis.

Guidelines for Use of Whole Exome Sequencing

The literature was perused to identify already existing clinical gene panels and to determine the relative advantages and disadvantages of WES compared to those panels. The information compiled in updating the gene lists was used to identify notable diagnostic features of different categories of NMDs, and this information was assembled into guidelines for the use of WES in patients with a suspected genetic NMD.

RESULTS

Diagnostic Yields are Improved with the Use of Broader Lists in Tandem with Focused Lists in Neuromuscular Disease Cases

In this cohort, a total of 32 patients were run on the myopathy diagnostic list and received a definitive response back, indicating either a (likely) pathogenic variant (positive), no pathogenic variant (negative), or a VUS (uncertain). Likewise, a total of 28 patients were run on the neuropathy diagnostic list and received a definitive result. When looking solely at the myopathy results, 27 patients were negative, 2 patients were uncertain, and 3 were positive, indicating a diagnostic yield of approximately 9%. For the neuropathy results, 23 patients were negative, 2 patients were uncertain, and 3 were positive, resulting in a diagnostic yield of 11%. Past analysis of NCGENES results have shown a roughly 12.7% diagnostic yield for patients enrolled in a neurological cohort, which includes NMD patients (Figure 2). More general data suggest that the diagnostic yield from WES is around 15-30%.[11]

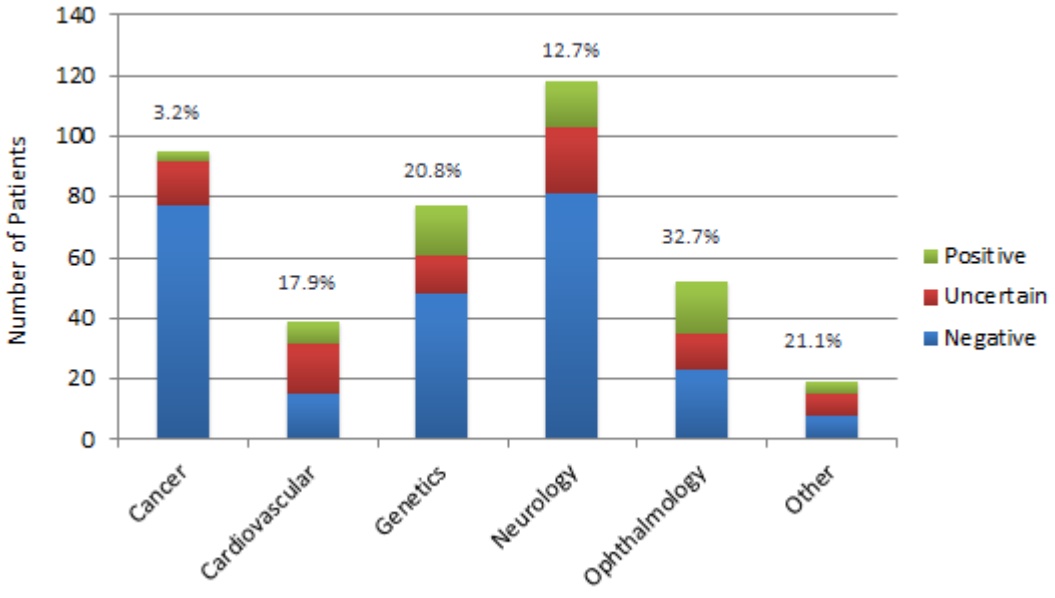


Figure 2. NCGENES diagnostic results and yields by disease category.

In this cohort, there were 23 patients who were run on both the myopathy and NMD lists, returning 4 positive, 3 uncertain, and 16 negative results. This resulted in a diagnostic yield of 17%. Of the four patients who received a positive result, one of them tested positive on both lists, while the other three tested positive on the NMD list only. Of the patients who had VUS, two reported VUS in both lists, while the other had a VUS in the NMD list and was negative on the myopathy list. From examining the clinical information of the four patients who received positive results, in patient NCG_00473, the variant that was found was a likely pathogenic (LP) missense variant in *MYH7*, which is consistent with Laing distal myopathy. Previously, the patient was suspected of having a limb-girdle muscular dystrophy. Patient NCG_00044 was found to have a frameshifting variant in COL9A3 consistent with autosomal dominant multiple epiphyseal dysplasia 3, a skeletal disorder characterized by early-onset short stature, waddling gait, and pain in the knees and other joints[12], providing a diagnosis when a significant amount of previous genetic testing had failed. NCG_00048 was found to have a pathogenic variant in GDAP1, which is associated with axonal Charcot-Marie-Tooth disease type 2K, a mild neuropathy phenotype.[13] The patient previously was tested for *SMN1* variants, leading to a diagnosis of a spinal muscular atrophy at age 3. Spinal muscular atrophies, which are characterized by degeneration of the anterior horn cells of the spinal cord, are the second most common lethal autosomal recessive disease in Caucasians, with the first being cystic fibrosis.[14] From the WES results, this patient received an updated diagnosis that overturned a long-standing previous one. Finally, NCG_00496 was diagnosed with a limb-girdle muscular dystrophy, and they were found to carry a homozygous frameshifting indel in *ANO5*, clarifying the specific type of limb-girdle muscular dystrophy.

There were 15 patients run on both the neuropathy and NMD lists, which returned 2 positive results, 4 uncertain results, and 9 negative results, indicating a diagnostic yield of 13%. Of the two patients who received positive results, in both instances, the positives came from both the neuropathy and NMD lists. Of the four patients who were found to have VUS, in one instance, the VUS came from both lists, whereas in the other three instances, the VUS came from the NMD list.

For both the myopathy and neuropathy cases, running the more specific list in tandem with the broader NMD list increased the diagnostic yield, though statistical significance was difficult to determine with the limited numbers. When the VUS results were considered in addition to the positive results, the increase in yield was greater. As comprehensive clinical information is not always available to the molecular analyst, a VUS result could be significant when subsequently analyzed by the ordering physician. Indeed, it may be warranted to further examine the VUS to clarify their effects, and cascade testing in affected family members is often performed. Regardless, for both combinations of lists, the positive and uncertain results predominantly came from the broader NMD list. The diagnostic yield data and interesting positive cases for the combined myopathy and NMD list are summarized in Figure 3 and Table 1, respectively.

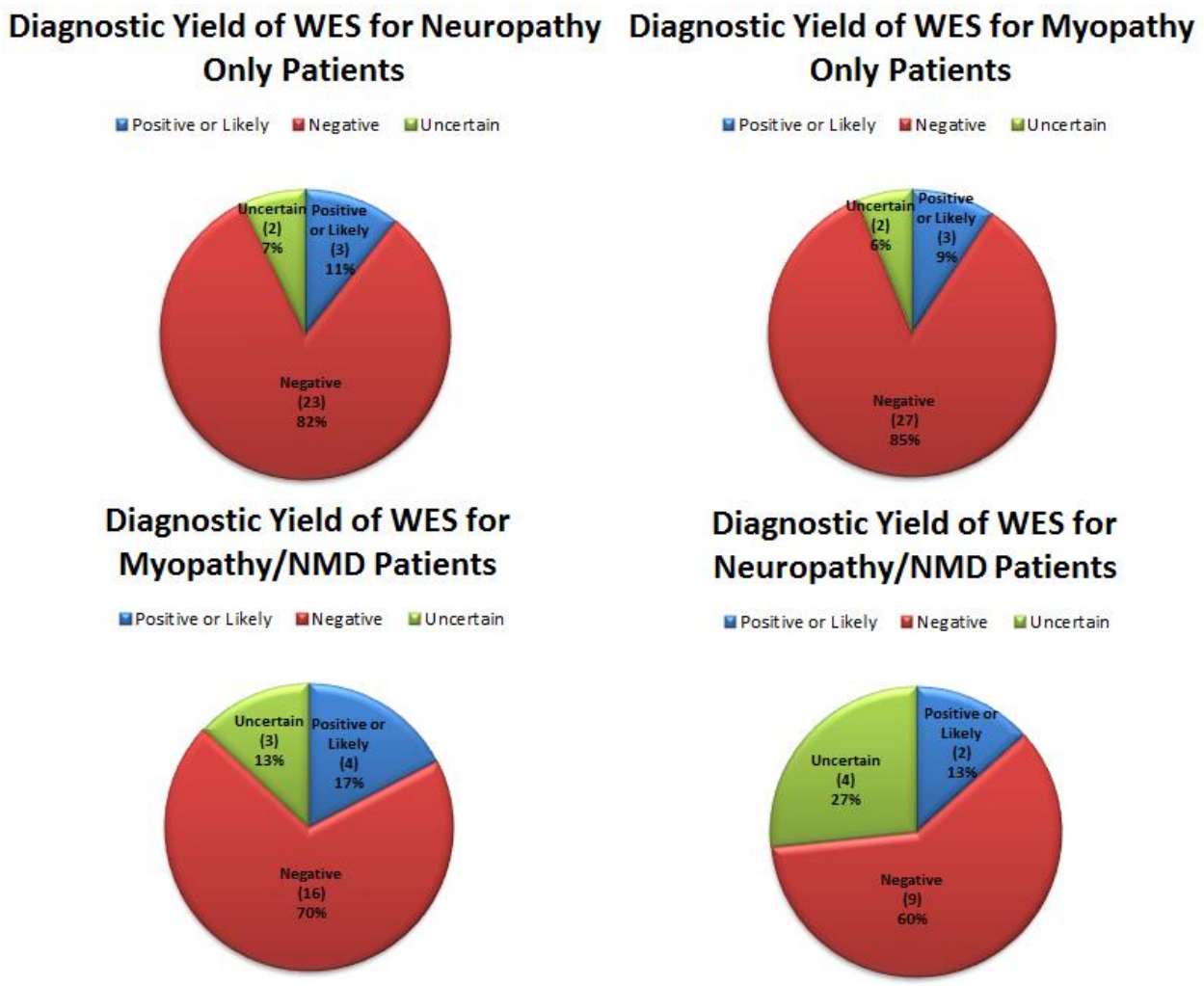


Figure 3. Diagnostic yields for NCGENES patients when run on only the neuropathy list (**upper-left**), only the myopathy list (**upper-right**), both the myopathy and NMD list (**bottom left**), and both the neuropathy and NMD list (**bottom right**).

ID	Working Clinical Diagnosis	Previous Genetic Testing	Positive or Likely Result	Clarification of Diagnosis
NCG_00473	Suspected LGMD	Unknown	*MYH7* MS LP consistent with Laing distal myopathy	Clarified diagnosis
NCG_00044		EDS I, II, IV; *MYOT* (LGMD 1a); Ullrich MD; LGMD 1B & C; FSHD	*COL9A3* frameshift for AD MED3	Provided diagnosis
NCG_00048	Spinal muscular atrophy diagnosis at age 3	*SMN1*	Pathogenic variant in *GDAP1* – CMT2K disease	Changed diagnosis
NCG_00496	Suspected LGMD	*FKRP*, DMD, FSHD	Homozygous *ANO5* frameshift indel for AR LGMD type 2L	Clarified diagnosis

Table 1. Clinical information of patients who received positive results from the combined myopathy and NMD lists. Italicized terms are gene names. **Working Clinical Diagnosis:** LGMD = limb-girdle muscular dystrophy. **Previous Genetic Testing:** EDS = Ehlers-Danlos syndrome, MD = muscular dystrophy, FSHD = facioscapulohumeral muscular dystrophy, DMD = Duchenne muscular dystrophy. **Positive or Likely Result:** MS = missense, LP = likely pathogenic, AD = autosomal dominant, MED3 = multiple epiphyseal dysplasia 3, CMT2K = Charcot-Marie-Tooth type 2K, AR = autosomal recessive.

Analysis of Negative Results Reveals a Potential Patient for an In-Depth Research Sweep

Neuromuscular disorders are an extremely broad category of inherited diseases that have a variety of inheritance patterns, including autosomal dominant or recessive. Consider the case of a causal variant that has either an autosomal dominant or autosomal recessive mode of inheritance. In the simplest case of the Hardy-Weinberg equilibrium of a single locus with two alleles with frequencies p and q, the equation is

$$p^2 + 2pq + q^2 = 1,$$

where p is the frequency of the major allele and q is the frequency of the minor allele. Assuming q is the frequency of the causal variant and the mode of inheritance is autosomal recessive, the expected overall prevalence of NMDs in the population based on the frequency of the causal variant would be $1{:}\frac{1}{q^2}$. Given the overall prevalence of NMDs, $\frac{1}{q^2}$ must be greater than at least 3,500.[1,2] So, q should have a frequency of at most 0.016, as $\frac{1}{0.016^2} = \frac{1}{2.56*10^{-4}} = 3906.25$, resulting in a calculated overall prevalence of 1 in 3,900. However, the initial frequency filter used to extract variants from the negative patient WES data was a minor allele frequency (MAF) of less than one percent from the Exome Aggregation Consortium (ExAC) database. This calculated frequency q is greater than the initial frequency limit applied to extract variants in the first place, so it would be too lenient to further filter the results. Assuming that the causal variant is inherited in an autosomal dominant fashion, the overall expected population prevalence of NMDs would be $1{:}\frac{1}{2pq+q^2}$. A reasonable value of q would then be dependent on the value of p.

Given that the initial frequency filter retrieved variants with an ExAC frequency of less than 1%, it would not be unreasonable to assume that p should have a frequency of at least 0.99. If that assumption is accepted, then a simple calculation using the quadratic formula shows that q should not exceed approximately 0.00014, which would give an overall prevalence of 1 in 3,600. These calculations assumed the overall prevalence of NMDs was around 1 in 3,500 (assuming the 1 in 3,000 prevalence would have allowed for higher MAF cutoffs, which is counterintuitive to further filtering variants). Given that the most common group of neuromuscular diseases – the Charcot-Marie-Tooth diseases or hereditary peripheral neuropathies – have an overall prevalence of roughly 1 in 2,500[15], setting the upper limit for q at 0.00014 appears reasonable. The overall prevalence is dominated by the more well-known (i.e.: more common) conditions, and if any patient truly has a "new" or "undiscovered" NMD, it would be expected to be rare. So, the upper limit for the MAF was determined to be 0.00014.

To identify a rare variant in a gene that could explain a particular disease phenotype, that variant should only appear in the patients with the disease and not in patients with unrelated illnesses. Frequencies were thus the primary and most important filter. When the initial frequency filter (only accepting variants with a MAF of less than one percent in ExAC) and the RDx list were applied to the WES data for patients who had received negative results, there were 4 truncating variants and 189 missense variants remaining within the cohort. All four truncating variants were rare enough to pass the subsequent limit set on the MAF, but in the process of examining the clinical notes on REDCap, it was found that three of the four patients were later found to have tested positive for variants in genes associated with an NMD that were not on the current diagnostic list. The remaining patient, NCG_00481, was found to be a promising candidate for a research sweep to specifically mine through all of their variants. This patient had

received mitochondrial genome sequencing, which was negative, and their microarray results were inconclusive. From the family history, the older brother had previously passed away of a similar suspected neuromuscular disorder, and there was a definite family history of symptoms on the maternal side, including hyperflexibility, myopathic features, and patchy mitochondrial swelling. This family history suggests something being inherited in an autosomal dominant manner. Taken together with the negative results from mitochondrial genome sequencing and WES, NCG_00481 would be a good candidate for further in-depth analysis by research sweeps.

For the missense variants, applying the MAF filter of 0.00014 reduced the number of variants from 189 to 85. Next, the Combined Annotation Dependent Deletion (CADD) score, a broad metric that integrates factors such as allelic diversity and experimentally-measured regulatory effects to score the deleteriousness of SNPs and indels, was examined. Typically, the standard in the discipline is to call variants with a score of 13 or greater to be considered deleterious. Applying this CADD filter then reduced the number of variants to 54. Each variant was inspected using the UCSC Genome Browser in order to check whether the variant was in a repetitive region, which would increase the likelihood of the variant being the result of sequencing miscalls or a sequencing artifact. Following inspection by the UCSC Genome Browser, the variants were examined for affected protein domains from UniProt and for published gene information and function from OMIM and GeneCards. This sweep has not yet identified a sufficiently rare variant with a plausible mechanism for contributing to an NMD, but those variants will still be retained for analysis by more sophisticated methods.

Curation of the Diagnostic Lists Adds Many More Genes and Changes the Structure of the Lists to Improve Clinical Use

Based on evaluations of the literature[2,3], which categorized myopathies and neuropathies under the broad umbrella of NMDs, it was decided that the NMD list should encompass all genes on the myopathy and neuropathy lists in addition to any potentially unique genes. This list structure would make the NMD list a broad-spectrum solution or method of analysis, which could be applied if the patient did not present with recognizable features of neuropathies or myopathies. For WES to be a useful research and diagnostic tool, the genes associated with the disorders of interest must be known, so the diagnostic lists must be continually updated to match new findings.

The initial lists were analyzed to examine the extent of gene overlap (numerical summary shown in Figure 4). Of the 80 genes on the neuropathy diagnostic list, all but 8 of them were also contained in the NMD diagnostic list. Likewise, 70 of the 99 genes on the myopathy diagnostic list were contained within the NMD diagnostic list. Of the 316 genes on the NMD diagnostic list, 177 were found solely on the NMD list. Each of these 177 genes was subsequently examined using the literature found on OMIM to see if it could also be classified as either a myopathy or a neuropathy gene. In total, 28 of those 177 genes were added to the myopathy list, and 18 were added to the neuropathy list. The summary of this survey of the literature, which reports the gene, the list it was added to, and the associated phenotypes, is displayed in Appendix 2.

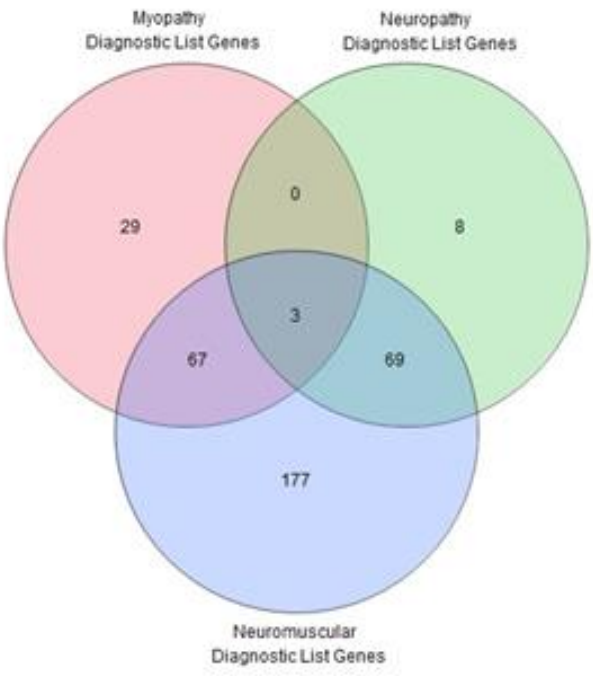


Figure 4. Distribution of genes among the three diagnostic lists. There is no overlap between the myopathy and neuropathy lists, and the NMD list encompasses large portions of both the neuropathy and myopathy lists.

In surveying the OMIM database of clinical synopses and phenotypes, 80 new genes were added to the myopathy list and 118 genes were added to the neuropathy list. In accordance with the decision made about the structure of the lists, all genes added to the myopathy and neuropathy lists were also added to the NMD list. In total, the NMD list expanded by 165 new genes, which reflected the fact that some of the genes added to the myopathy and neuropathy lists were already on the NMD list. The three new diagnostic lists were reexamined to determine the level of overlap, with a numerical summary shown in Figure 5. The organization of these three news lists matched our expectations, with the NMD list encompassing all of the neuropathy and myopathy lists, and the myopathy and neuropathy lists being largely unique individuals. The updated diagnostic lists for myopathy, neuropathy, and NMDs lists are displayed as tables in Supplemental 2.

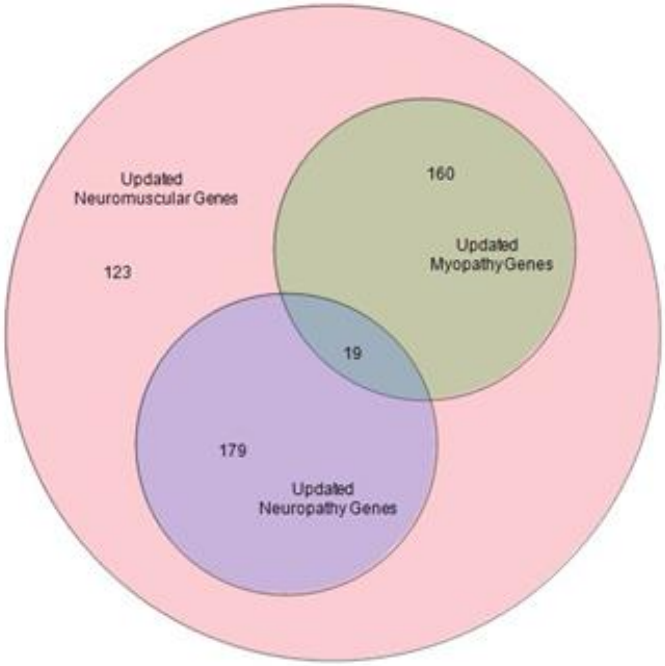


Figure 5. Distribution of genes among the three diagnostic lists after addition of new genes discovered through manual surveys of the literature.

Guidelines for Incorporating WES into a Clinical Workflow for Patients with Suspected Neuromuscular Disorders

We hypothesized that knowledge of any distinguishing features of types of NMDs would narrow down the possible genes that could contain a causal variant, improving the efficiency of analyzing WES data and improving the likelihood that a patient would get a positive result. WES typically has good coverage of approximately 85%-90% of the exons in the genome, leaving inadequate coverage for about 1,000 to 2,000 genes.[5,16] As a result, negative and uncertain results should not be taken as an absolute rejection. This emphasizes the importance of clinical phenotypic information in analyzing WES output, so knowledge of characteristic features for different types of NMDs would be beneficial for the physician in selecting an appropriate genetic test (which may or may not include WES) and for the molecular analyst in identifying and verifying causal variants. Curating the literature[17] has identified several clinical features often

indicative of a congenital myopathy, as well as distinguishing muscle biopsy features, which are summarized in Tables 2.1-2.2. These features were often found to be particular to a specific genetic diagnosis, but considering that the aim of the study is to aid the clinician in the use of WES, I considered it more reasonable to provide the clinician with categories of details to look for so that they would be able to provide more detailed information to the molecular analyst, improving the efficiency and confidence of variant calling.

Clinical Feature	Age Range
Facial weakness	Newborns and children less than 2 years
Opthalmoplegia (extraocular muscle weakness)	
Ptosis (drooping of eyelids)	
Facial dysmorphisms (long face or head relative to width, high arched palate)	
Bulbar weakness (sucking or swallowing)	
Significant respiratory involvement at birth	
Severe congenital hypotonia (“floppy infant”)	
Predominant axial hypotonia	
Orthopedic deformities	
Dislocation of hip	
Club feet	
Fetal akinesia or severe arthrogryposis (joint contracture)	
Scoliosis	Older children from 2 years to adolescence
Rigid spine	
Cardiomyopathy	
Foot drop/pes cavus	
Malignant hyperthermia	
Respiratory and axial involvement disproportionate to skeletal muscle weakness	

Table 2.1 Clinical features found to be suggestive of a congenital myopathy, divided by age ranges where those features might be most useful in a diagnostic context.

Structural Defect	Description
Rods	Cytoplasmic and occasionally intranuclear bodies indicative of nemaline myopathies; Rod-like or ovoid, with or without attached filaments; Variable in distribution and number; Stained red by Gomori trichrome technique; Electron microscopy reveals lattice structure similar to Z-line; Immunohistochemistry reveals proteins similar to Z lines, especially α-actinin
Cores	Variable appearance, but classically devoid of mitochondria and therefore oxidative enzyme activity; May be peripheral or central; May be multiple per fiber; Extends down a significant length of the fibers
Central nuclei	Large bodies that are spaced at intervals down the fiber and occupy large volumes of the fiber; Number of nuclei may increase with age; In neonates, often associated

	with dark stained centers that have a pale peripheral halo with oxidative enzyme staining
Caps	Peripheral, well-demarcated areas eosinophilic with a haematoxylin and eosin stain; No ATPase activity and myosin staining; Positively label for actin and α-actinin; Electron microscopy shows focal, peripheral areas of myofilaments oriented in multiple directions, usually with attached thin filaments and thickened Z-lines
Minicores	Multiple focal areas devoid of oxidative enzyme activity, and appear as such in longitudinal sections; Electron microscopy is a useful tool for identifying focal areas of myofibrillar disruption
Hyaline bodies	Focal areas that are seen as granular, basophilic zones with a haematoxylin and eosin stain; Positive for ATPase activity and demonstrate slow myosin staining; Electron microscopy shows them as granular areas
Necklace fibers	Feature a clear ring or loop of oxidative enzyme staining internally within the fiber and not attached to the sarcolemma

Table 2.2 Pathological muscle biopsy features found to be suggestive of a congenital myopathy.

Like the inherited myopathies, the inherited neuropathies can be very broadly divided into those in which the neuropathy is the primary part of the disease and those in which the neuropathy is part of a more generalized, often multisystem disorder (i.e.: a syndromic neuropathy.) As a result, I classified some of the distinguishing characteristics of inherited neuropathies into broad categories and related them to some disease phenotypes. The literature findings[17-20] of these categories are summarized in Table 3.

Clinical Feature	Associated Phenotypes
Uniform slowing of nerve conduction velocities	Charcot-Marie-Tooth 1 (demyelinating - motor conduction velocities of the median or ulnar nerve less than 38 m/s); Dejerine-Sottas disease; metachromatic leukodystrophy; Krabbe's disease; Cockayne's disease
Asymmetric or multifocal slowing of nerve conduction velocities	Charcot-Marie-Tooth X-linked; Refsum's disease; Pelizeus-Merzbacher disease; adrenomyeloneuropathy
Sensory dysfunction (auditory and visual, of limbs, etc.)	Hereditary sensory and autonomic neuropathies; familial amyloid polyneuropathies; neuropathies associated with mitochondrial disorders; autosomal dominant Charcot-Marie-Tooth (axonal); autosomal recessive Charcot-Marie-Tooth
Distal limb involvement	Autosomal dominant Charcot-Marie-Tooth 1; peripheral neuropathies; distal hereditary motor neuropathies (muscle weakness, but typically no sensory loss); hereditary sensory and autonomic neuropathies

Proximal limb involvement	Autosomal recessive Charcot-Marie-Tooth; peripheral neuropathies

Table 3. Groups of clinical features suggestive of neuropathies, with some associated phenotypes. Due to the phenotypic heterogeneity of inherited neuropathies, these categories of symptoms may be found in other neuropathies or NMDs, but they are often key signs of a neuropathy phenotype.

These features provide clinical evidence of a particular disease phenotype, but the importance of accurate genetic diagnosis cannot be overstated. Therefore, the clinician must make a decision about the appropriate testing to use. While WES has been proven to be a powerful diagnostic approach, there are limitations that include a large list of generated variants, incomplete coverage with capture libraries, and difficulty detecting variants resulting from repeat expansion, copy-number variation, epigenetic alterations, and some others (Table 4).[1,9,21] Therefore, WES would not be the test of choice for diseases associated with these types of variants or that have existing clinical panels that cover a large majority of the known genes. In these cases, WES would either not be able to detect causal variants, or another clinical panel could be used to return results for less cost or in less time. I found GeneTests (www.genetests.org) to be an excellent up-to-date resource for finding the availability and cost of genetic testing in both diagnostic and research laboratories, and a summary of some available tests for NMDs is included in Table 5. Based on gene information, availability of testing, and clinical symptoms, we devised a decision-making tool for incorporating WES in a clinical setting for patients with suspected neuromuscular disease (Appendix 3).

Variant Type	Associated Phenotypes
Repetitive DNA (including trinucleotide repeats)	Fragile X syndrome, Huntington's disease, **myotonic dystrophy 2**, **X-linked spinal and bulbar muscular atrophy, oculopharyngeal muscular dystrophy**
Copy-number variants	DiGeorge syndrome (22q11.2 deletion syndrome), **Charcot-Marie-Tooth disease type 1**, **facioscapulohumeral muscular dystrophy**
Structural variants (including chromosomal translocations and inversions; large deletions,	**Duchenne/Becker muscular dystrophy, spinal muscular atrophy 1**

duplications, and rearrangements)	
Long indel variants	Resistance to HIV infection
Aneuploidy	Down's syndrome, Turner syndrome
Epigenetic alterations	Prader-Willi syndrome, Beckwith-Wiedemann syndrome

Table 4. Variant types currently not well-detected or undetected by exome sequencing, with examples of associated phenotypes. NMD phenotypes are written in bold.

Disorders	Laboratory	Method	Genes Tested	Turnaround Time and Cost
NMDs	Emory Molecular Genetics Laboratory	N/A	ACTA1, AMPD1, AMPD3, ANO5, CAPN3, CAV3, COL6A1, COL6A2, COL6A3, DES, DMD, DYSF, EMD, FKRP, FKTN, GAA, GNE, ISPD, ITGA7, LAMA2, LARGE, LMNA, MYOT, NEB, PLEC, PMM2, POMGNT1, POMT1, POMT2, PYGM, RYR1, RYR2, SEPN1, SGCA, SGCB, SGCD, SGCE, SGCG, SIL1, TCAP, TNNI2, TNNT1, TPM2, TPM3, TRIM32, TTN	12-13 weeks; Contact lab
NMDs	GeneDx	NGS, Deletion and duplication (copy-number variation)	ACTA1, ANO5, ATP2A1, B3GALNT2, BAG3, BIN1, CACNA1S, CAPN3, CAV3, CFL2, CHKB, CLCN1, CNTN1, COL6A1, COL6A2, COL6A3, CRYAB, DAG1, DES, DMD, DNAJB6, DNM2, DPM1, DPM2, DPM3, DYNC1H1, DYSF, EMD, FHL1, FKRP, FKTN, FLCN, GAA, GMPPB, GNE, IGHMBP2, ISPD, ITGA7, KBTBD13, KLHL40, LAMA2, LAMP2, LARGE, LDB3, LMNA, MEGF10, MTM1, MYH7, MYOT, NEB, PLEC, PLEKHG5, POMGNT1, POMT1, POMT2, RYR1, SCN4A, SEPN1, SGCA, SGCB, SGCD, SGCG, SIL1, SYNE1, TCAP, TMEM5, TNNI2, TNNT1, TNPO3, TPM2, TPM3, TRIM32, TRPV4, TTN, UBA1, VRK1	8-9 weeks; Contact lab
Congenital myopathies	PreventionGenetics	NGS	ACTA1, BIN1, CCDC78, CFL2, CNTN1, DNM2, KBTBD13, KLHL40, KLHL41, MTM1, MYF6, MYH7, NEB, RYR1, SEPN1, STAC3, TNNT1, TPM2, TPM3	4-6 weeks; $2,590.00
Congenital myopathies and muscular dystrophies	GeneDx	NGS, Deletion and duplication (copy-number variation)	ACTA1, CFL2, CHKB, COL6A1, COL6A2, COL6A3, FKRP, FKTN, ITGA7, KBTBD13, LAMA2, LMNA, MEGF10, NEB, RYR1, SEPN1, SYNE1, TNNT1, TPM2, TPM3	8-9 weeks; Contact lab
Charcot-Marie-Tooth (Full Panel)	Invitae	NGS, Deletion and duplication (copy-number variation)	AARS, BSCL2, DNM2, EGR2, FGD4, FIG4, GARS, GDAP1, GJB1, HSPB1, HSPB8, KIF1B, LITAF, LMNA, MED25, MFN2, MPZ, MTMR2, NDRG1, NEFL, PMP22, PRPS1, PRX, RAB7A, SBF2, SH3TC2, TRPV4, YARS	2-3 weeks; $1,500.00
Hereditary neuropathies	Emory Molecular Genetics Laboratory	N/A	ADCK3, ALDH3A2, APTX, ATL1, ATM, C10orf2, CACNB4, CTDP1, DCTN1, DNM2, EGR2, FGD4, FGF14, FXN, GAN, GARS,	12-13 weeks; Contact lab

			GDAP1, GJB1, HOXD10, HSPB1, HSPB8, IKBKAP, ITPR1, KCNA1, KCNC3, KIAA0196, KIF1B, KIF5A, L1CAM, LITAF, MFN2, MPZ, MRE11A, MTMR2, MTTP, NDRG1, NEFL, NGF, NIPA1, PEX7, PHYH, PLP1, PMP22, PNPLA6, POLG, PRKCG, PRPS1, PRX, RAB7A, REEP1, SACS, SBF2, SH3TC2, SIL1, SLC12A6, SLC1A3, SPAST, SPG11, SPG20, SPG21, SPG7, SPTBN2, SPTLC1, TDP1, TTBK2, TTPA, WNK1, YARS, ZFYVE26, ZFYVE27	

Table 5. Some genetic panels available for neuromuscular disorders, with the tested disorders, covered genes, turnaround time, lab location, and cost. Information taken from GeneTests (www.genetests.org).

DISCUSSION

Since its inception, whole exome sequencing and indeed, massively parallel sequencing strategies as a whole, have proven to be very powerful in identifying causal variants in disease. However, as with any new technology, time and effort must be dedicated to ensure proper implementation and best practices. The objective of this study then, was to tackle the question of what must be done to make the most use of whole exome sequencing in patients with neuromuscular disorders, a broad and heterogeneous group. Though this study is preliminary, the analysis has identified several points that will be important in guiding the use of WES for this unique and diverse subset of patients.

First and foremost is the importance of making sure that the diagnostic lists are as complete and up-to-date as possible. WES of human samples that have 100x or more coverage (the average number of reads representing a given nucleotide in the reconstructed sequence) generate around 20,000-30,000 single nucleotide variants (SNV) and indel calls, according to some estimates.[23] The diagnostic list is a method to prioritize variants that are relevant to the disease, by allowing the analyst to consider only those variants present in known disease genes. The lists are also important for identifying potentially novel disease variants. For example, a novel rare non-synonymous SNV found in a gene known to cause a phenotype is more likely to

prove to be causative as well.[5] Discovering new disease variants or genes also better illustrates genetic functions and pathways by providing an opportunity to investigate deviations from normal activity. In light of just how much the diagnostic lists were expanded, it may be more fruitful to wait for the results of a reanalysis of the WES data before further attempting to identify novel disease genes or variants. So, a significant amount of time and effort was spent scouring the literature to both add new known disease genes to the lists and to ensure that the genes already on the list were accurate. Curating the lists to ensure they stay current is a task that must be continuously done, as more information is constantly discovered about variants that supports or rejects its role in disease.

A major challenge of implementing WES in a clinical environment is determining which patients, and which diseases, are most amenable to diagnosis using this method. As such, assessing the diagnostic yield, or how often patients receive a result that can explain their phenotype, is paramount to addressing this obstacle. It is clear, both from the literature and previous NCGENES data, that the diagnostic yield varies for different genetic conditions. Data from NCGENES indicates a past 12-13% yield in patients with neurological conditions, while literature values suggest an overall diagnostic yield of 15-30%[11,24,25], and some evidence suggesting a yield of around 20-30% for neurological disorders.[25] As shown, the diagnostic yields we obtained for our neuropathy, neuromuscular, and NMD patients was lower than those values, but those results are likely understated. First is the matter of the VUS, which as defined, could actually contribute to the disease phenotype. While speculative, if the VUS was a heterozygous variant, it could be possible that the patient's phenotype has an oligogenic mode of inheritance, so there could be another heterozygous variant in a different, though functionally-related, gene, and together, they result in the patient's phenotype. A probably more realistic

explanation is that the VUS simply should be further characterized to try to ascertain its deleteriousness. Second, and more obvious, is the fact that the initial prioritization of variants was incomplete because the diagnostic lists used to streamline the variant search were incomplete. Evaluation of the literature doubled the size of the myopathy and neuropathy lists and increased the size of the NMD list by 50%. Most likely, a reanalysis of those patients' WES data using the new lists to cultivate the variant search will result in significantly higher diagnostic yields, and this analysis is ongoing.

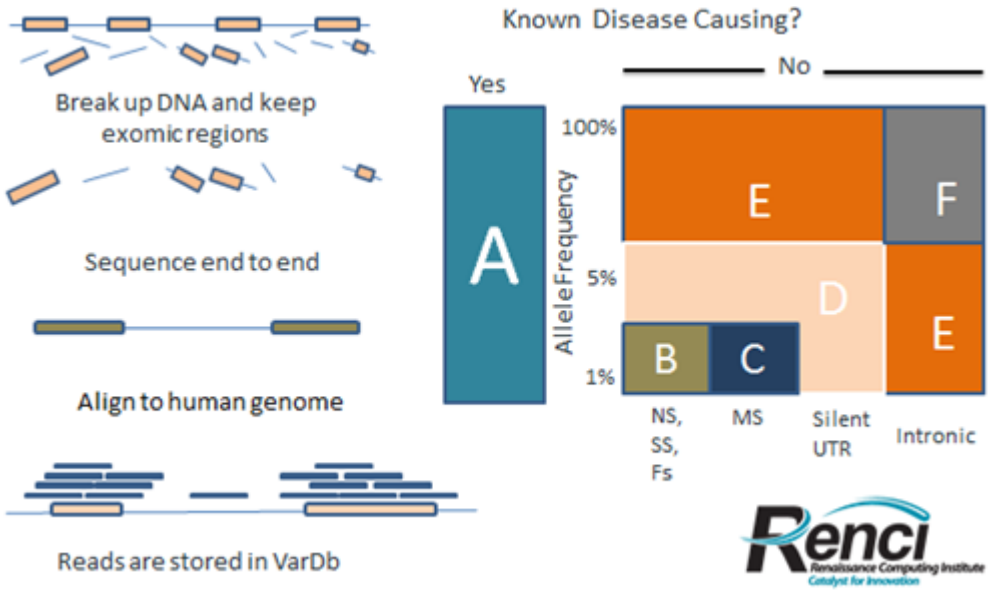
Another question related to the use of the diagnostic lists is deciding which lists should be applied in the variant search. This is largely a classic question of efficiency, for a broader diagnostic list would be more likely to return positive variants at the cost of time, effort, and computational power. The initial data analyzed here is perhaps too limited in scale to make broad assertions (though that is usually a problem when dealing with rare diseases), but the information gleaned from the literature combined with the data gives us at least some confidence. We found that patients who were run on the broad NMD list in tandem with a more focused list had greater diagnostic yields, and their positive results tended to come from the broad lists. This combination of lists also returned more VUS (which again, typically came from the NMD list), which could warrant further characterization and lead to a deleterious variant that explains that phenotype. Now consider the four positive cases in the combination myopathy/NMD list. Each of those patients had either no or inconclusive testing in the past, limiting the amount of information available in the present to build off of for a diagnosis. Even so, WES was able to provide them with a (clarified) diagnosis or even change an old one. In addition, analysis of the literature has indicated several clinical features and neuromuscular examination results that are predictive of a specific genetic diagnosis while at the same time emphasizing the need for genetic testing in

order to confirm it. It is, as they say, the only way to be sure. So, these findings at least suggest that when patients do not present with indicative symptoms (especially muscle or nerve biopsy and nerve conduction velocity results), it is advantageous to use the broader, more encompassing diagnostic gene list. The method of construction of the updated myopathy, neuropathy, and NMD lists, in which the NMD list encompasses all myopathy and neuropathy genes while also containing its own unique genes, reflects this philosophy. I anticipate that reanalyzing the patients using the updated diagnostic lists will not only return more positive results, it will also provide more of our own empirical data to connect certain symptoms, test results, or family histories to a particular genetic diagnosis.

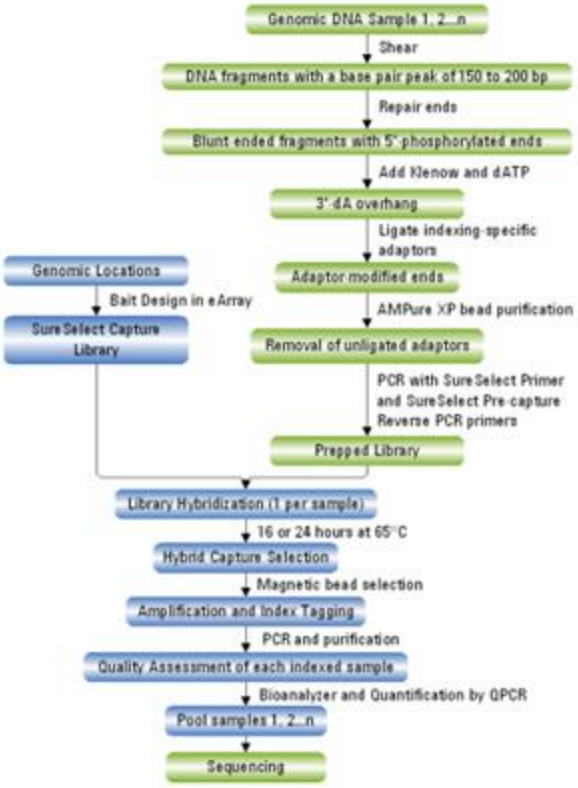
The ultimate goal of this study, and a major goal of NCGENES as a whole, is to establish evidence-based guidelines for the use of whole exome sequencing for diagnosis. Based on the analysis of the diagnostic yields and the work done to curate the lists, it seems our framework is a reasonable one. WES is currently indicated for detection of rare variants in patients with a suspected highly-penetrant Mendelian disorder, after traditional single-gene diagnostic methods have not resulted in a diagnosis or multigene approaches would be too expensive.[9] Our framework reflects that, calling for the use of WES in situations where the cost or technical limitations would not prohibit it. If a clinical panel already exists, the sequencing coverage of the panel compared to WES is a key consideration. For medical diagnosis, a mean depth of coverage of 100x or more is desirable, as it greatly reduces the difficulty in identifying and filtering out false positive variant calls.[5] If, for example, there are 20,000-30,000 variant calls[23], even a 1% false positive rate introduces hundreds of false calls, complicating diagnostic applications. Typically, WES platforms have 50x coverage, though this can be pushed up to 100x. Clinical gene panels are almost guaranteed to have significantly higher coverage, so it is advantageous in

terms of cost, analysis workload, and accuracy to use a clinical panel if one that covers enough genes exists. Especially important in this framework is what the clinician needs to report when submitting samples for WES. By essentially being as detailed as possible about clinical findings, family histories, and test results, the clinician provides important information that improves the confidence of variant calling and analysis. Future steps that would improve these guidelines would include using the reanalyzed patient data to provide a statistical analysis of the predictive power of different clinical items towards a specific genetic diagnosis and providing a more in-depth cost and coverage analysis of available clinical panels compared to WES.

APPENDIX 1: NCGENES WHOLE-EXOME SEQUENCING WORKFLOW



Appendix 1.1. The WES workflow and variant calling pipeline for NCGENES. The library preparation and sequencing steps are further detailed in Appendix 1.2. The sequencing done is paired-end sequencing. When the variants are called, they are given a designation based on their predicted or known pathogenicity. The "A" designation is for variants that are known to cause disease. Designations "B" through "F" are in descending order of predicted pathogenicity. The smaller the allele frequency and the more potentially disruptive a variant is, the higher it is ranked and the more likely it is to be pathogenic.



Appendix 1.2. The overall sequencing library preparation workflow for the SureSelect^{XT} Target Enrichment System Kit for Illumina Multiplexed Sequencing. Paired-end sequencing is used.

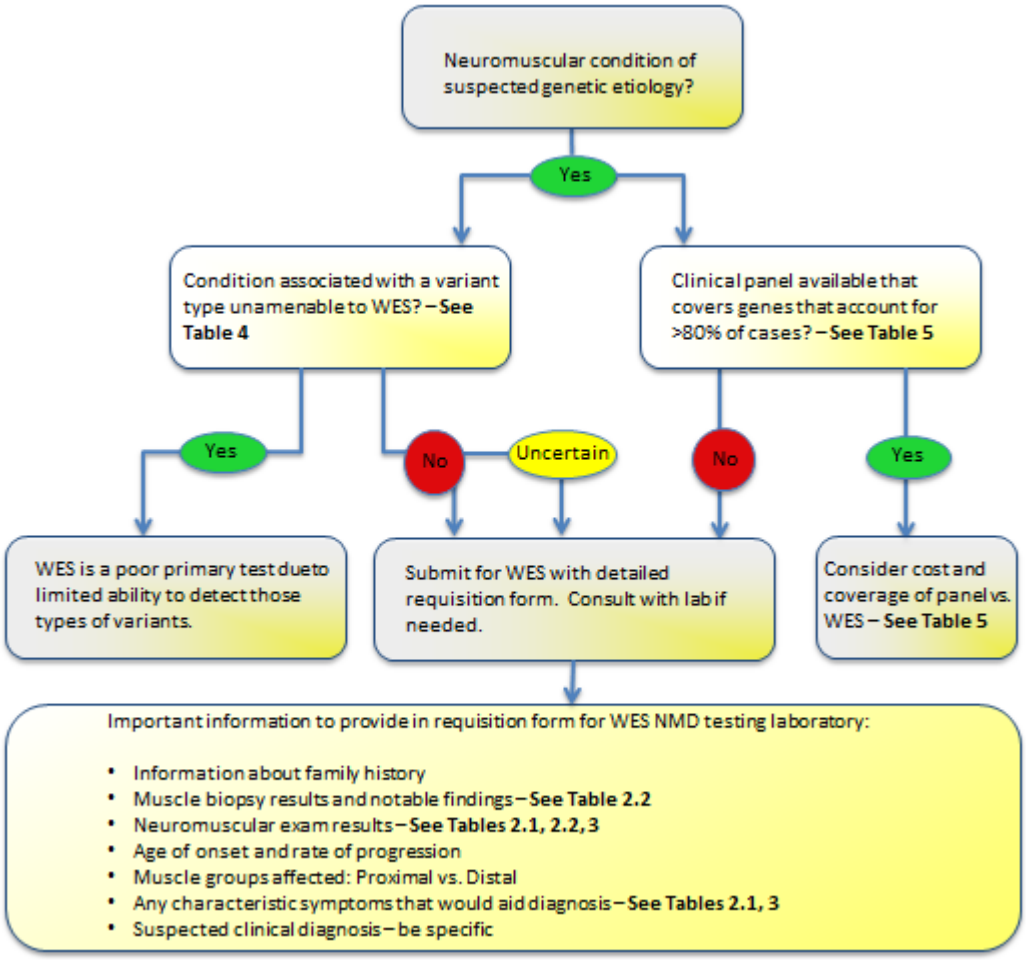
APPENDIX 2: CURATION OF THE NEUROMUSCULAR DISEASE DIAGNOSTIC LIST

Genes Added to the Myopathy List		Genes Added to the Neuropathy List	
Gene Name (HGNC)	**Disease Phenotype**	**Gene Name (HGNC)**	**Disease Phenotype**
FXN	Friedrich ataxia (with or without retained reflexes)	FXN	Friedrich ataxia (with or without retained reflexes)
MYH14	Autosomal dominant deafness 4A; Peripheral neuropathy, myopathy, hoarseness, and hearing loss	MYH14	Autosomal dominant deafness 4A; Peripheral neuropathy, myopathy, hoarseness, and hearing loss
TTPA	Ataxia (Friedrich-like) with isolated vitamin E deficiency	TTPA	Ataxia (Friedrich-like) with isolated vitamin E deficiency
ABHD5	Chanarin-Dorfman syndrome	GJC2	AR spastic paraplegia 44; Hereditary lymphedema 1C; Hypomyelinating leukodystrophy 2
CCDC78	Centronuclear myopathy 4	INF2	Dominant intermediate Charcot-Marie-Tooth E; Glomerulosclerosis, focal segmental 5
COL6A1	Bethlem myopathy; Ullrich congenital muscular dystrophy	LRSAM1	Axonal Charcot-Marie-Tooth 2P
COL6A2	Bethlem myopathy; Ullrich congenital muscular dystrophy; congenital myosclerosis	NF1	Neurofibromatosis type 1
COL6A3	Bethlem myopathy; Ullrich congenital muscular dystrophy	PDSS1	Primary coenzyme deficiency 2

COL9A3	Multiple epiphyseal dysplasia 3; Multiple epiphyseal dysplasia with myopathy	PDYN	Spinocerebellar ataxia 23
DYSF	Miyoshi muscular dystrophy 1; Limb-girdle muscular dystrophy type 2B; Distal myopathy with anterior tibial onset	REEP1	Autosomal dominant spastic paraplegia 31; Distal hereditary motor neuronopathy type VB
FKBP14	Ehlers-Danlos syndrome with progressive kyphoscoliosis, myopathy, and hearing loss	SCN4A	Paramyotonia congenita, Congenital myotonia (acetazolamide-responsive); Acetazolamide-responsive myasthenic syndrome; Hyper and hypokalemic periodic paralysis
FKRP	Muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies; congenital with or without mental retardation; limb-girdle)	SETX	Juvenile amyotrophic lateral sclerosis 4; Autosomal recessive spinocerebellar ataxia 1
FLNC	Distal myopathy 4; Myofibrillar myopathy 5	SPG11	Autosomal recessive spastic paraplegia 1
LDB3	Dilated cardiomyopathy 1C; Left Ventricular noncompaction 3, with or without dilated cardiomyopathy; Myofibrillar myopathy 4	SPTLC2	Hereditary sensory and autonomic neuropathy, type IC
MEGF10	Myopathy, areflexia, respiratory distress, and dysphagia (early onset)	TRPV4	Brachyolima type 3; Familial digital arthropathy-brachydactyly; Hereditary motor and sensory neuropathy type IIc; Metatropic dysplasia; Parastremmatic dwarfism; Scapuloperoneal spinal muscular atrophy; Spondyloepiphyseal dysplasia Maroteaux type; Distal congenital nonprogressive spinal muscular atrophy; Kozlowski type spondylometaphyseal dysplasia
MTMR14	Modifier of autosomal centronuclear myopathy	TUBB3	Congenital fibrosis of extraocular muscles type 3A; Complex cortical dysplasia
MTTP	Myopathy; Merff syndrome; Susceptibility to Parkinson disease	ZFYVE26	Autosomal recessive spastic paraplegia 15
MYBPC3	Dilated cardiomyopathy 1MM; Familial hypertrophic cardiomyopathy 4; Left ventricular noncompaction 10	IGHMBP2	Axonal Charcot-Marie-Tooth 2S; Distal hereditary motor neuronopathy type VI
MYF6	Centronuclear myopathy 3		
MYH3	Distal arthrogryposis type 2A and 2B		
NEB	Autosomal recessive nemaline myopathy 2		
PLEC	Epidermolysis bullosa simplex (Ogna type; with pyloric atresia; with epidermolysis bullosa simplex); Limb-girdle muscular dystrophy type 2Q		
SGCD	Dilated cardiomyopathy 1L; Limb-girdle muscular dystrophy type 2F		
SIL1	Marinesco-Sjorgen syndrome		
TCAP	Limb-girdle muscular dystrophy type 2G; Dilated cardiomyopathy 1N		
TNNT3	Distal arthrogryposis type 2B		
TPM2	CAP myopathy; Autosomal dominant nemaline myopathy 4; Distal arthrogryposis type 2B; Distal arthrogryposis multiplex congenita type 1		
VCP	Amyotrophic lateral sclerosis 14 with or without frontotemporal dementia; Inclusion body myopathy with early-onset Paget disease and frontotemporal dementia		

Appendix 2. Genes on the NMD diagnostic list that were added to the myopathy or neuropathy lists upon inspection using the OMIM database. Many genes were associated with more than one disease phenotype, but not all of those disease phenotypes were associated with a neuropathy or myopathy.

APPENDIX 3: INCORPORATING WHOLE EXOME SEQUENCING INTO A CLINICAL SETTING FOR PATIENTS WITH NEUROMUSCULAR DISEASE



Appendix 3. A possible method of incorporating whole exome sequencing (WES) in a clinical setting for patients with suspected neuromuscular disease.

SUPPLEMENTAL MATERIALS

Supplemental materials have been included with this submission as additional PDF files.

REFERENCES

1. Jobanputra V, Naini A. New Developments in Molecular Diagnosis of Neuromuscular Disorders. *J. Neurol. Disord. Stroke* **2013**; *1*: 1003-1004.

2. Emery AE. Population Frequencies of Inherited Neuromuscular Diseases – A World Survey. *Neuromuscul. Disord.* **1991**; *1*: 19-29.

3. Laing NG. Genetics of Neuromuscular Disorders. *CRC. Cr. Rev. Cl. Lab. Sc.* **2012**; *49*: 33-48.

4. Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively Parallel Sequencing and Rare Disease. *Hum. Mol. Genet.* **2010**; *19*:R119-124.

5. Marian AJ. Sequencing Your Genome: What Does it Mean? *Methodist Debakey Cardiovasc. J.* **2014**; *10*: 3-6.

6. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. Exome Sequencing Identifies the Cause of a Mendelian Disorder. *Nat. Genet.* **2010**; *42*: 30-35.

7. Teer JK, Mullikin JC. Exome Sequencing: the Sweet Spot Before Whole Genomes. *Hum. Mol. Genet.* **2010**; *19*: 145-151.

8. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The Next-Generation Sequencing Revolution and its Impact on Genomics. *Cell* **2013**; *155*: 27-38.

9. Biesecker LG, Green RC. Diagnostic Clinical Genome and Exome Sequencing. *N. Engl. J. Med.* **2014**; *370*: 2418-2425.

10. Bao R, Huang L, Andrade J, Tan W, Kibbe WA, Jiang H, Feng G. Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. *Cancer Inform.* **2014**; *13*: 67-82.

11. Volk A, Conboy E, Wical B, Patterson M, Kirmani S. Whole-Exome Sequencing in the Clinic: Lessons from Six Consecutive Cases from the Clinician's Perspective. *Mol. Syndromol.* **2015**; *6*: 23-31.

12. Muragaki Y, Mariman ECM., van Beersum SEC, Perala M., van Mourik JBA, Warman ML, Olsen BR, Hamel BCJ. A Mutation in the Gene Encoding the Alpha-2 Chain of the Fibril-Associated Collagen IX, COL9A2, Causes Multiple Epiphyseal Dysplasia (EDM2). *Nature Genet.* **1996**; *12*: 103-105.

13. Zimon M, Baets J, Fabrizi GM, Jaakola E, Kabzinska D, Pilch J, Schindler AB, Cornblath DR, Fischbeck KH, Auer-Grumbach M, et al. Dominant GDAP1 Mutations Cause Predominantly Mild CMT Phenotypes. *Neurology* **2011**; *77*: 540-548.

14. Wirth B. An Update of the Mutation Spectrum of the Survival Motor Neuron Gene (SMN1) in Autosomal Recessive Spinal Muscular Atrophy (SMA). *Hum. Mutat.* **2000**; *15*: 228-237.

15. Grandis M, Shy ME. Current Therapy for Charcot-Marie-Tooth Disease. *Curr. Treat. Options Neurol.* **2005**; *7*: 23-31.

16. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, et al. Exome Sequencing and the Genetic Basis of Complex Traits. *Nat. Genet.* **2012**; *44*: 623-630.

17. Lewis RA, Sumner AJ, Shy ME. Electrophysiological Features of Inherited Demyelinating Neuropathies: A Reappraisal in the Era of Molecular Diagnosis. *Muscle Nerve* **2000**; *23*: 1472-1487.

18. Axelrod, FB, Hilz MJ. Inherited Autonomic Neuropathies. *Semin. Neurol.* **2003**; *23*: 381-390.

19. Reilly, MM. Classification and Diagnosis of the Inherited Neuropathies. *Ann. Indian. Acad. Neurol.* **2009**; *12*: 80-88

20. Botez, SA, Herrmann, DN. Sensory Neuropathies, from Symptoms to Treatment. *Curr. Opin. Neurol.* **2010**; *23*: 502-508.

21. Lieberman AP, Fischbeck KH. Triplet Repeat Expansion in Neuromuscular Disease. *Muscle Nerve* **2000**; *23*: 843-850.

22. Rabbani B, Tekin M, Mahdieh N. The Promise of Whole-Exome Sequencing in Medical Genetics. *J. Hum. Genet.* **2014**; *59*: 5-15.

23. O'Rawe J, Jiang T, Sun G, et al. Low Concordance of Multiple Variant-Calling Pipelines: Practical Implications for Exome and Genome Sequencing. *Genome Med.* **2013**; *5*: 28.

24. Levenson D. Whole-Exome Sequencing Emerges as Clinical Diagnostic Tool. *Am. J. Med. Genet.* **2014**; *164*: ix-x.

25. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, et al. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA*. **2014**; *312*: 1870-1879.